

NPS55-82-028

# NAVAL POSTGRADUATE SCHOOL

Monterey, California



USE OF VOICE RECOGNITION EQUIPMENT  
WITH STENOGRAPHER MASKS

by

G. K. Poock  
N. D. Schwalm  
E. F. Roland

October 1982

Approved for public release; distribution unlimited.

Prepared for:  
9th Infantry Division  
Fort Lewis, WA 98433

DUDLEY KNOX LIBRARY  
NAVAL POSTGRADUATE SCHOOL  
MONTEREY, CA 93943-5101

NAVAL POSTGRADUATE SCHOOL  
Monterey, California

Rear Admiral J. J. Ekelund  
Superintendent

D. A. Schrady  
Provost

Reproduction of all or part of this report is authorized.

This report was prepared by:

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

## REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS  
BEFORE COMPLETING FORM

1. REPORT NUMBER NPS55-82-028		2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) USE OF VOICE RECOGNITION EQUIPMENT WITH STENOGRAPHER MASKS		5. TYPE OF REPORT & PERIOD COVERED Technical	
		6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s) G. K. Poock N. D. Schwalm E. F. Roland		8. CONTRACT OR GRANT NUMBER(s)	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, CA 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS MIPR TB-024	
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, CA 93940		12. REPORT DATE October 1982	
14. MONITORING AGENCY NAME & ADDRESS (If different from Controlling Office) 9th Infantry Division Fort Lewis, WA 98433		13. NUMBER OF PAGES 53	
		15. SECURITY CLASS. (of this report) UNCLASSIFIED	
		15a. DECLASSIFICATION/ DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.			
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, If different from Report)			
18. SUPPLEMENTARY NOTES			
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) VTAG Voice Recognition Automatic Speech Recognition Voice Input/Output			
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This research examined the effectiveness of entering command to voice recognition equipment when speaking through a stenographer's mask. The results indicate it is feasible and promising. -			



## TABLE OF CONTENTS

	<u>Page</u>
LIST OF FIGURES	iii
LIST OF TABLES	iv
EXECUTIVE SUMMARY	v
1. INTRODUCTION	1-1
1.1 Background	1-1
1.2 Problem	1-2
1.3 Objective	1-2
2. METHOD	2-1
2.1 Subjects	2-1
2.2 Apparatus	2-1
2.3 Experimental Design	2-2
2.4 Procedures	2-5
2.4.1 Training	2-5
2.4.2 Testing	2-6
2.5 Independent and Dependent Variables	2-6
3. RESULTS	3-1
3.1 Overview	3-1
3.2 Total Errors	3-2
3.3 Nonrecognitions (Rejections)	3-6
3.4 Misrecognitions	3-6
3.5 Experience with Masks and Microphones	3-11
4. DISCUSSION	4-1
4.1 Total Errors	4-1
4.2 Nonrecognitions	4-2
4.3 Misrecognitions	4-2
4.4 Experience with Masks and Microphones	4-3
5. CONCLUSIONS	5-1
6. REFERENCES	6-1

## TABLE OF CONTENTS (CONT.)

APPENDIX A - LIST OF UTTERANCES

APPENDIX B - QUESTIONNAIRE

## LIST OF FIGURES

<u>FIGURE NO.</u>	<u>DESCRIPTION</u>	<u>PAGE</u>
2-1	Illustration of Subject Using Voice Recognition Equipment Through the Stenographer Mask	2-3
2-2	Summary of Experimental Design	2-4
3-1	Total Error Rates by Mask Conditions by Trials	3-5
3-2	Misrecognition Error Rates by Mask Conditions by Trials	3-10
3-3	Misrecognition Error Rates by Levels of Mask Experience Mask Conditions	3-18
3-4	Misrecognition Error Rates by Levels of Microphone Experience by Mask Conditions	3-19

## LIST OF TABLES

<u>TABLE NO.</u>	<u>DESCRIPTION</u>	<u>PAGE</u>
3-1	Analysis of Variance Summary Table for Total Errors	3-3
3-2	Mean Total Error Rates for Mask Condition by Trials	3-4
3-3	Mean Percent Nonrecognitions for Mask Conditions by Trials	3-7
3-4	Analysis of Variance Summary Table for Misrecognitions	3-8
3-5	Mean Misrecognition Rates for Mask Conditions by Trials	3-9
3-6	Pearson Product-Moment Correlations Between Experience with Masks and Microphones and the Dependent Measures	3-12
3-7	Analysis of Variance Summary Table for Misrecognitions with Mask Experience as the Between Group Variable	3-14
3-8	Analysis of Variance Summary Table for Misrecognitions with Microphone Experience as the Between Group Variable	3-15
3-9	Mean Misrecognition Error Rates for Levels of Mask Experience by Mask Conditions	3-16
3-10	Mean Misrecognition Error Rates for Levels of Microphone Experience by Mask Conditions	3-17

## EXECUTIVE SUMMARY

The purpose of this experiment was to determine if the high accuracy rate of current voice recognition systems would be reduced significantly if speakers were required to enter utterances through a mask, as opposed to the "boom" microphone used with most conventional voice recognition systems. It is conceivable that voice recognition equipment may, in fact, be used in the near future in multi-purpose, high-activity command, control, and communication (C<sup>3</sup>) centers, where several speakers will undoubtedly need to operate voice recognition devices at the same time.

The findings suggest that no significant increase in non-recognitions (e.g., errors where the system rejects the input and says, in effect, "I don't understand you, say it again") is evident while using a mask. Misrecognitions (i.e., errors where the system accepts the input but mistakes it for a different input) do increase significantly under masked conditions. However, the data also indicate that prior experience with speaking into masks or microphones may be a significant moderator of this relationship; subjects that reported having had little or no experience speaking into masks or microphones showed significantly more misrecognition errors than those that reported having some or considerable experience speaking into masks or microphones. Moreover, the data indicate that, when using masks, those subjects that reported having had experience with speaking into masks and microphones (e.g., pilots, communicators) displayed misrecognition error rates still statistically different from but much more comparable to the error rates displayed by subjects under no-mask conditions.

Since misrecognitions, as defined earlier, may be potentially a more critical type of error, it is suggested that training individuals on how to speak into masks or microphones should reduce significantly the number of misrecognitions that may occur under masked conditions. It is concluded

that current voice recognition equipment may be used effectively under masked conditions without practically significant performance decrement (as compared to no-mask conditions), provided that users are adequately trained. Further research should investigate the amount of training required to achieve optimal accuracy of currently available voice recognition equipment in situations where operators may be required to use masks. It is also clear that the costs of such training must be kept relatively low so that the current benefits of using "voice" as opposed to conventional input modes are maintained.

## 1. INTRODUCTION

### 1.1 Background

In recent years, voice technology has developed to the extent that basic systems have now been used successfully in several industrial and military applications. With constant improvements being made in the capabilities of voice recognition systems, their use in a wider variety of settings is already being contemplated.

One such setting is that of the forward observer (FO) in the Army's TACFIRE system. The FO currently uses a keyboard to relay formatted information back to the control 0 console of the TACFIRE system which is usually located in a large mobile van. The FO also uses voice communications in his tasks. Given the proper equipment configuration, it might be possible to use voice recognition/input equipment at the FO position to verbally enter information and relay it to the TACFIRE van.

Another setting which could be considered as a candidate for the use of voice recognition/input is at the artillery control console in the TACFIRE van itself. This console is activated through the use of manual typing into a keyboard which controls artillery direction and other items of information. This van is really a command and control center for a variety of actions. Given the proper equipment configuration, it may also be possible to use voice recognition/input in the command center atmosphere of the TACFIRE van itself.

### 1.2 Problem

The problem which may exist in both examples above is a preponderance of environmental noises around the voice recognition user (the speaker). In the case of the FO, environmental noises may be quite loud and of the impact

type at times. In the case of a voice input operator in the TACFIRE van, other people in the van talking or yelling may cause problems for an operator trying to enter voice commands.

One could possibly solve both of these noise problems by blocking out the surrounding noise if the operator talked into some type of mask with a microphone in it. Such a mask does currently exist and is known as a stenographer's mask for use in court rooms where a stenographer can input voice transactions without being heard by others in the room. This same mask is being tested by the Army for use by personnel operating close to enemy positions. It is intended to muffle the voice while engaged in radio communications.

Could such a mask be used to input commands through a voice recognition system and still maintain high levels of recognition accuracy by the voice recognizer?

Specifically, does the impressive accuracy rate ascribed to currently available voice recognition equipment suffer significantly if the user is required to enter utterances to the system through a mask, as opposed to the conventional "boom" microphone mounted on a headset?

Relatively recent research (Elster, 1980) showed that background noise (including speech) did not interfere significantly with voice recognition accuracy. This is encouraging, since it implies that "voice" would be effective in  $C^3$  centers where much background activity may be anticipated. Little research, however, has been done on the effectiveness of voice in larger installations where several speakers, each operating a separate recognizer, may be required to make inputs simultaneously. It is conceivable that, under those conditions, the speakers or operators themselves might become confused by each other's speech, thus perhaps increasing *input* errors. This could also be the case in command briefings, where a speaker

may be required to communicate with others not in the immediate area; having to raise one's voice to get another's attention could interfere with ongoing activities and cause confusion. Thus, two kinds of situations (recognizer inaccuracy and speaker confusion) could produce the same results--inappropriate output by the "voice" system.

### 1.3 Objective

The specific objective of the present research was to assess empirically the accuracy with which a currently available voice recognition system would interpret utterances that were input through stenographer's masks as compared to the conventional "boom" microphone input device normally worn on an operator's head.

Specific research is currently being conducted using Army gas masks also, which would be another type of mask worn for protection in a nuclear, biological and chemical warfare environment. The results of the gas mask study will be reported soon in another report.

(Note: The results of the current study with stenographer's masks also has direct technology transfer to many types of command briefs or morning briefs in all military services. An operator could be sitting right in the briefing room and listening to the conversations to know what situation displays or other graphic information needed to be displayed. By speaking into a stenographer's mask, the operator could be using voice recognition to bring up displays, etc., and it would all happen silently without disrupting the briefing.)



## 2. METHOD

### 2.1 Subjects

Thirty-six subjects (32 males, 4 females) originally participated in the study. All subjects were volunteers recruited from curriculums at the Naval Postgraduate School in Monterey, California. It should be noted that due to the lengthy period over which the present study was conducted, one of the T600 voice recognition systems was needed for other purposes on a large enough number of occasions so as to make it unavailable to the researchers on a consistent basis. Therefore, the analyses that follow are based on only half (18) of the 36 subjects that began the experiment. Although this may theoretically have reduced the power of the statistical tests used, the author feels that the within-groups design coupled with the elaborate counterbalancing scheme used still allows for reliable interpretation of the results.

Thus, the study was essentially carried out using 18 subjects (14 males, 4 females). Their ages ranged from 25 to 36 years, with a median age of 31 years.

### 2.2 Apparatus

Two Threshold Technology model T600 voice recognition devices were used in this study. Each of these devices was capable of handling 256 two-second voice utterances; 100 utterances were used in the present investigation. A list of these utterances is contained in Appendix A. For more details on the operation of voice recognition equipment see Poock (1980).

Three input devices were used in the experiment. The first was the conventional Shure model SM10 "boom" microphone (mounted on a headset), which is supplied as standard equipment with the T600. The second input

device was a stenographer's mask (STENOMASK) manufactured by Talk, Incorporated of Westbury, N.Y. This contained a Shure model 99L86LF microphone, supplied as standard equipment by the manufacturer. The third input device was a STENOMASK identical to that mentioned above. However, this mask was modified to contain the same SM10 microphone implanted in the same housing as the standard STENOMASK microphone. That is, the device was identical to the standard STENOMASK except for the microphone itself; the difference between the two masks was visually undetectable. Inclusion of the STENOMASK with the SM10 microphone would enable the researchers to attribute differences in recognition accuracy to the mask itself, rather than to any particular microphone. Figure 2-1 illustrates a subject using the T600 under masked conditions.

### 2.3 Experimental Design

A 6x3x6 mixed design with repeated measures on two factors was employed in this experiment. The first factor, order of mask use, was the between variable, and was comprised of the 6 orders in which all three masks could be used by each subject; subjects were nested within this variable such that six subjects received one of the six possible "mask" orders. This counterbalancing scheme was adopted to control for any effects that order of use may have contributed to the results. "Mask" condition (N = No Mask, O = Original Mask, S = Shure Mask) was a three-level, within group variable with each subject performing under each of the three "mask" conditions. Each subject also performed 6 trials with each mask, making trials the second within group variable with 6 levels. A summary of the experimental design appears in Figure 2-2.



FIGURE 2-1.  
SUBJECT USING THE T60R MASK

TRIALS	NO MASK (N)						ORIGINAL MASK (O)						SHURE MASK (S)					
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
ORDER OF MASK USE	S-N-O		$s_1$ →															
			⋮															
			$s_3$ →															
	S-O-N		$s_4$ →															
			⋮															
			$s_6$ →															
S-N-O	N-S-O		$s_7$ →															
			⋮															
			$s_9$ →															
	N-O-S		$s_{10}$ →															
			⋮															
			$s_{12}$ →															
O-S-N	O-S-N		$s_{13}$ →															
			⋮															
			$s_{15}$ →															
	O-N-S		$s_{16}$ →															
			⋮															
			$s_{18}$ →															

FIGURE 2-2  
SUMMARY OF EXPERIMENTAL DESIGN

## 2.4 Procedure

2.4.1 Training. The term "training," as used in discussions of voice recognition studies, refers to the process by which the speaker makes known to the recognizer the characteristics of his particular speech patterns for all the utterances he will be using. For the T600, this training procedure consists of entering 10 passes of each utterance (10x100 or 1,000 utterances in this study) into the voice recognizer. The recognizer automatically enters these utterances into its "memory," and matches any subsequent utterances of the same vocabulary (in testing) with those in memory. Ideally, these subsequent utterances are matched with those in memory and the result is a correct response output on a CRT. In cases where the recognizer can not make this match, a nonrecognition or rejection occurs, and this results in a "beep" from the recognizer; in effect, the machine is saying "I don't understand that utterance--please say it again." Occasionally, however, the recognizer "thinks" it has matched an utterance with one in memory, but the match is incorrect. In this case, an incorrect response is output on the CRT, constituting what is known as a "misrecognition." Thus, two types of errors are possible: nonrecognitions (or rejections) and misrecognitions (or misinterpretations) of an utterance.

For training, each subject spoke 10 passes of each of 100 utterances into the voice recognizer (total = 1,000 utterances). It was necessary to do this once for each mask condition under which subjects served. This procedure took approximately one hour for each training session. Due to the relatively large number of subjects used in this study, it was necessary for half of the subjects to come in on Monday and half on Tuesday on each of three weeks (one week per mask condition). Since half the subjects came in on one of those days and half on the other, any variability in training performance was also theoretically controlled. Subjects trained the system on Monday (or Tuesday) for all 3 training

sessions. Immediately after training, subjects made at least two passes of the entire 100 word vocabulary (essentially a test session) to identify any problems in training of any particular utterance. Where the system produced correct responses on those two passes, the utterance was considered adequately trained. If errors occurred (of either type) a third pass was made. If less than two of three passes of any utterance was correct, that utterance was retrained.

2.4.2 Testing. After training, subjects tested the system. Each subject was scheduled to make two passes through the entire vocabulary list on each of three successive days. These testing sessions were administered on Wednesday, Thursday, and Friday of the same week in which training took place. Thus, a total of six testing trials were run for each subject under each "mask" condition. In this way, subjects were able to complete training and testing of one mask condition within one week. The experiment ran for a total of three weeks, with one mask condition being run each week.

## 2.5 Independent and Dependent Variables

The independent variable in this study was "mask" condition: No Mask, where subjects trained and tested the system using the conventional "boom" microphone; and original Mask, where subjects trained and tested the stenomask containing the standard microphone supplied by the manufacturer; and Shure Mask, where subjects trained and tested the stenomask containing the Shure SM10 microphone.

The dependent variables in this study were nonrecognitions (or rejections), misrecognitions, and total errors, which was a linear combination of nonrecognitions and misrecognitions.

At the conclusion of the experiment, each subject was asked to fill out a questionnaire designed to measure certain attitudes and experience variables that the researchers felt might affect performance. This questionnaire appears in Appendix B.



### 3. RESULTS

#### 3.1 Overview

This section describes the results of the present study. All analyses were performed using the SPSS (Nie, Hull, Jenkins, Steinbrenner and Bent, 1975) and BMDP (Brown, Engelman, Frane, Hill, Jennrich and Toporek, 1981) statistical packages. All repeated measures analyses of variance procedures were performed using the arcsin transformation of raw data to stabilize the variance of the error terms (Neter and Wasserman, 1974). The mean error rates that appear in the figures, however, are untransformed. All *a posteriori* tests for significance between pairs of means were performed using the Scheffe procedures described in Bruning and Kintz (1977).

As defined earlier, nonrecognitions and misrecognitions by the voice recognition system may have distinctly different implications in an applied setting. To take an extreme example, in a weapons deployment activity, it would be far more desirable for the system to respond to an input error by nonrecognition (a "beep"), where the speaker is essentially told that he should repeat the input (or correct it), than for the system to misinterpret the input and to carry out some incorrect (and perhaps critical) command in error. Thus, it was considered essential to determine the effects of the independent variables on nonrecognitions and misrecognitions separately, as well as on total number of errors (nonrecognitions + misrecognitions).

Section 3.2 presents the data for total number of errors. Section 3.3 presents the results of analyses done on nonrecognitions or rejections, while Section 3.4 presents the results of analyses done on misrecognitions. Finally, Section 3.5 presents the results on misrecognitions in light of subjects' past experience speaking into masks and microphones.

### 3.2 Total Errors

Table 3-1 presents the analysis of variance summary table for total errors (Nonrecognitions + Misrecognitions). Significant main effects of mask condition ( $F = 12.92$ ,  $p < .01$ ) and trials ( $F = 3.18$ ,  $p < .01$ ) are evident. Order of mask use was not a significant effect, nor were there any significant interactions. Mean error rates (in percent) are shown in Table 3-2, and the main effect of mask condition and trials are portrayed graphically in Figure 3-1.

With regard to the main effect of mask condition, a Scheffé test for significance between pairs of means was performed to determine between which pairs of means the significant differences lie. The results of this test indicated that significant differences existed between the no mask condition and both original and shure mask conditions. The differences between the original and shure mask conditions was not significant.

A review of Figure 3-1 indicates that performance deteriorated over trials, most saliently for the original mask condition, and somewhat for the no mask condition.

Although one might think of fatigue as an explanation of this trials effect, this seems to be implausible, since only two test trials were run on any given day and each lasted less than 5 minutes. It is possible that because the later trials took place toward the end of a school week, subjects were not as alert as they were in the middle of the week when the earlier test trials took place. The author therefore suggests that the trials effect evident in Figure 3-1 may be spurious rather than systematic in nature.

TABLE 3-1.  
ANALYSIS OF VARIANCE SUMMARY TABLE FOR TOTAL ERRORS.

Source of Variance	df	MS	F
Order (O)	5	0.27	0.82
Error	12	0.32	-
Mask Condition (M)	2	1.49	12.92*
M x O	10	0.10	0.87
Error	24	0.11	-
Trials (T)	5	0.06	3.18*
T x O	25	0.02	0.96
Error	60	0.02	-
M x T	10	0.02	1.00
M x T x O	50	0.02	1.09
Error	120	0.02	-

\* p < .01

TABLE 3-2.  
MEAN TOTAL ERROR RATES (IN PERCENT) FOR MASK CONDITIONS BY TRIALS

MASK CONDITIONS

	NO MASK	ORIGINAL MASK	SHURE MASK	$\bar{x}$ TRIALS
TRIAL 1	1.56	3.89	5.39	3.61
TRIAL 2	1.61	4.00	5.44	3.68
TRIAL 3	1.56	4.28	5.22	3.69
TRIAL 4	1.72	5.50	5.17	4.13
TRIAL 5	2.22	7.94	4.94	5.03
TRIAL 6	2.11	6.83	5.33	4.76
$\bar{x}$ MASKS	1.80	5.41	5.25	GRAND $\bar{x}$ 4.15

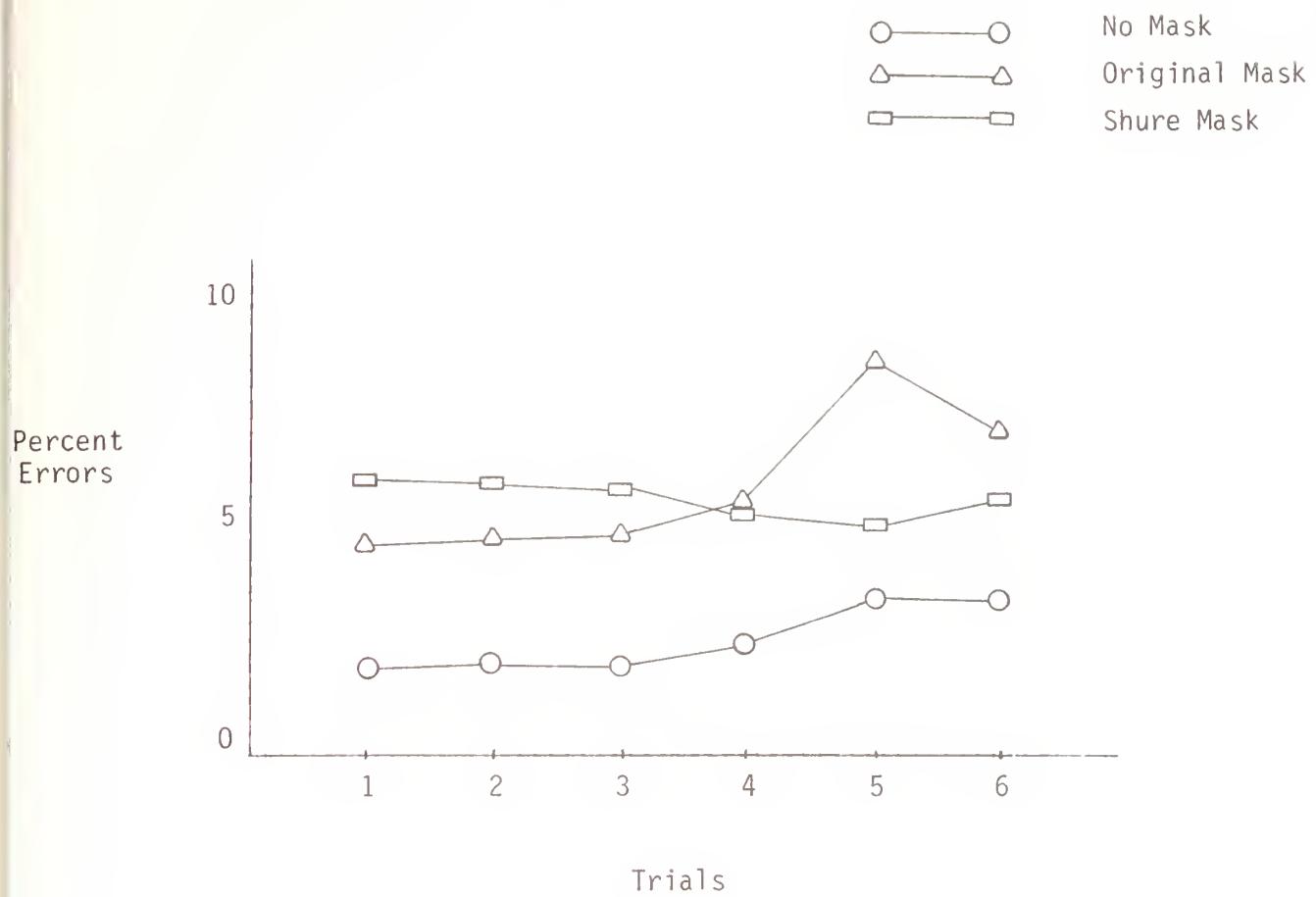


FIGURE 3-1.  
TOTAL ERROR RATES BY MASK CONDITIONS BY TRIALS

### 3.3 Nonrecognitions (Rejections)

An analysis of variance was performed on the nonrecognitions alone to determine the effects, if any, of the independent variables. No significant effects of order of mask use, mask condition, or trials were found, nor were there any significant interactions. Table 3-3 presents the percent nonrecognitions by trials by mask conditions.

### 3.4 Misrecognitions

As was done for nonrecognitions, an analysis of variance was performed on the misrecognitions alone, to determine the effects of the independent variables. Table 3-4 presents the analysis of variance summary table for misrecognitions.

Significant main effects of mask condition ( $F = 12.57, p < .01$ ) and trials ( $F = 3.50, p < .01$ ) are evident. Order of mask use was not found to be a significant effect, nor were there any significant interactions. Mean misrecognition rates (in percent) are shown in Table 3-5, and the main effects of mask condition and trials are portrayed graphically in Figure 3-2.

With regard to the main effect of mask condition, a Scheffé test for significance between pairs of means was performed to determine between which pairs of means the significant differences lie. The results of this test indicated that significant differences existed between the no mask condition and both original and shure mask conditions. The differences between the original and shure mask conditions were not significant.

A review of Figure 3-2 indicates that performance deteriorated over trials, most saliently for the original mask condition and somewhat for the no mask condition. As in the case of total errors, the author is not clear as to the reason for this deterioration, and maintains that this effect is probably not a systematic effect, especially because it is not evident with regard to the other mask condition.

TABLE 3-3.

MEAN PERCENT NONRECOGNITIONS BY TRIAL BY MASK CONDITION.

## MASK CONDITION

	NO MASK	ORIGINAL MASK	SHURE MASK	$\bar{x}$ TRIALS
TRIAL 1	0.67	0.11	0.78	0.52
TRIAL 2	0.50	0.17	0.83	0.50
TRIAL 3	0.44	0.72	0.72	0.63
TRIAL 4	0.56	0.50	0.83	0.63
TRIAL 5	0.50	1.44	1.05	0.99
TRIAL 6	0.28	1.78	0.83	0.96
$\bar{x}$ MASKS	0.49	0.79	0.84	GRAND $\bar{x}$ 0.71

TABLE 3-4.  
ANALYSIS OF VARIANCE SUMMARY TABLE FOR MISRECOGNITIONS.

Source of Variance	df	MS	F
Order (O)	5	0.25	0.72
Error	12	0.34	-
Mask Condition (M)	2	1.42	12.57*
M x O	10	0.09	0.76
Error	24	0.11	-
Trials (T)	5	0.05	3.50*
T x O	25	0.02	1.15
Error	60	0.02	-
M x T	10	0.02	0.85
M x T x O	50	0.02	1.24
Error	120	0.02	-

\* p < .01

TABLE 3-5.  
MEAN MISRECOGNITION RATES (IN PERCENT)  
FOR MASK CONDITIONS BY TRIALS.

MASK CONDITIONS

	NO MASK	ORIGINAL MASK	SHURE MASK	$\bar{x}$ TRIALS
TRIAL 1	0.89	3.77	4.61	3.09
TRIAL 2	1.11	3.83	4.61	3.18
TRIAL 3	1.11	3.56	4.50	3.06
TRIAL 4	1.17	5.00	4.33	3.50
TRIAL 5	1.72	6.50	3.88	4.03
TRIAL 6	1.83	5.06	4.50	3.80
$\bar{x}$ MASKS	1.31	4.62	4.41	GRAND $\bar{x}$ 3.44

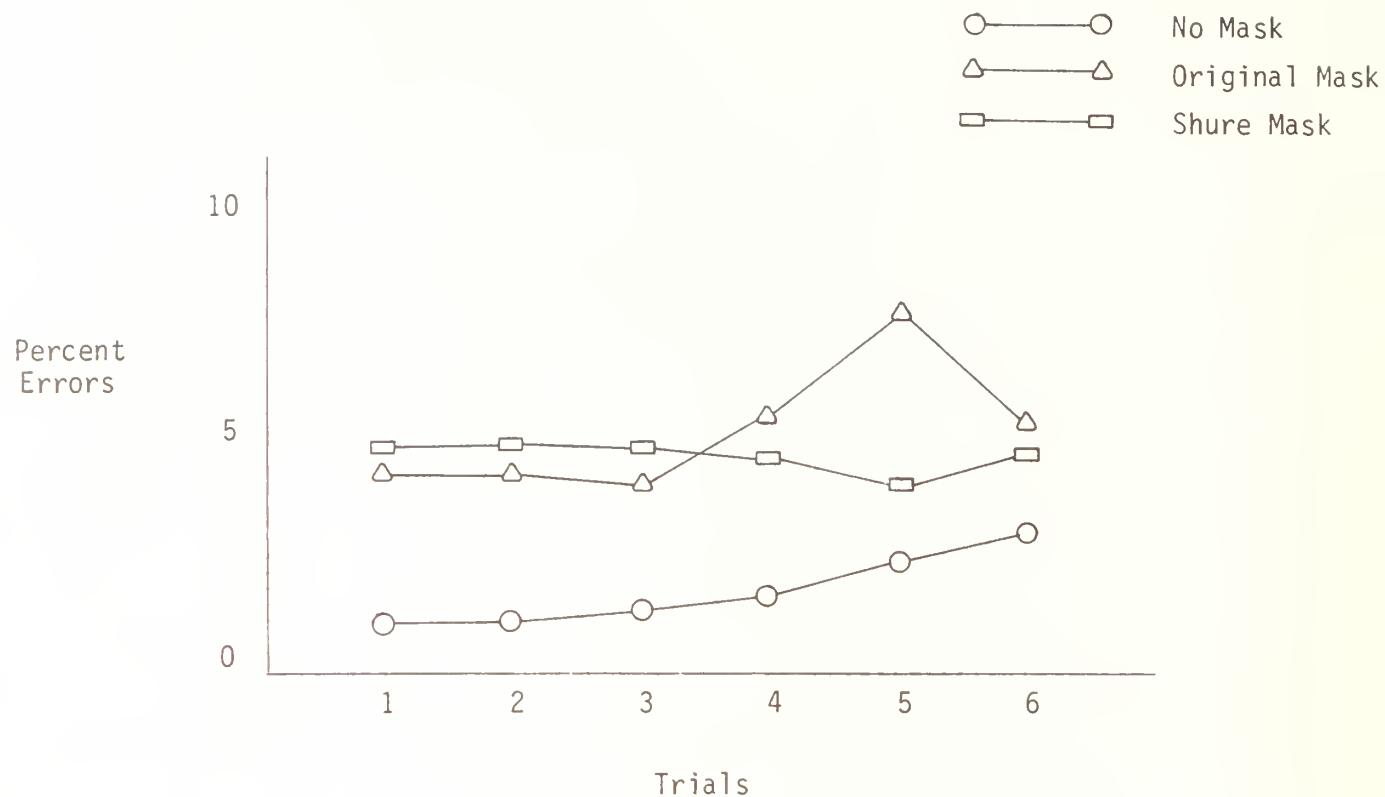


FIGURE 3-2.  
MISRECOGNITION ERROR RATES BY MASK CONDITIONS BY TRIALS

A review of Figures 3-1 and 3-2 indicates a strong similarity in the nature of the total error and misrecognition data. This, coupled with the absence of significant differences in nonrecognitions, makes it apparent that the real differences in error rates due to mask conditions are reflected primarily in misrecognitions.

### 3.5 Experience with Masks and Microphones

It was noted earlier that, at the conclusion of the last testing session, a questionnaire was administered to the subjects that was designed to assess the extent of their experience with speaking into masks or microphones. These data were subjected to a series of analyses to determine their moderating effect on misrecognition errors.

The first step in determining whether experience with masks or microphones was related to the dependent measures was to perform a Pearson Product-Moment correlation procedure on the data. The results of those correlations appear in Table 3-6 for each mask condition. The correlations across all mask conditions were: misrecognitions with mask experience:  $r_{xy} = -0.55$ ,  $p < .01$ ; misrecognitions with microphone experience:  $r_{xy} = -0.53$ ,  $p < .02$ . Overall, nonrecognitions did not correlate significantly with either mask or microphone experience. The size and direction of these significant correlations suggests that the more experience subjects had with masks or microphones (primarily with masks), the fewer misrecognition errors were made. These results prompted the author to perform a series of analyses of variance on the misrecognition data to determine the exact nature of the experience effects.

Subjects were divided into three groups: Group 1 was comprised of all subjects that scored three or below on the seven-point experience scales (for both masks and microphones) and were called the "low" experience groups; Group 2 was comprised of all subjects that scored four on the scales, and

TABLE 3-6.  
PEARSON PRODUCT MOMENT CORRELATIONS BETWEEN EXPERIENCE  
WITH MASKS AND MICROPHONES AND THE DEPENDENT MEASURES

TYPE OF ERROR

MASK CONDITION -	MISRECOGNITIONS			NONRECOGNITIONS		
	NO MASK	ORIGINAL MASK	SHURE MASK	NO MASK	ORIGINAL MASK	SHURE MASK
Experience With Masks	-0.41**	-0.43**	-0.54*	-0.41**	-0.25	-0.19
Experience with Microphones	-0.22	-0.37	-0.59*	-0.28	-0.30	-0.05

\* p < .05

\*\* p < .10

were called the "intermediate" experience groups; Group 3 was comprised of all subjects that scored five and above on the scales, and were called the "high" experience groups. These groups comprised the between variable in two analyses of variance procedures identical to the ones performed previously (where order of mask use was a six-level between group variable).

It should be noted that, with regard to the breakdown of subjects by experience with microphones, only two groups (high and low experience) emerged; there were no subjects who described themselves as having only "some" (intermediate) experience with microphones. Thus, the analysis of variance procedure for microphone experience included only a two-level between group variable instead of a three-level between group variable, as in the case of mask experience.

The analysis of variance summary tables appear in Tables 3-7 and 3-8 for mask and microphone experience respectively. Review of these tables makes it apparent that experience is a significant moderator of misrecognition errors in both cases (as suggested by the correlation coefficients reported earlier). Mean misrecognition rates (in percent) are shown in Tables 3-9 and 3-10 for mask and microphone experience variables respectively. Figures 3-3 and 3-4 portray graphically the percent of misrecognition errors by mask condition by mask and microphone experience levels respectively. (Note that due to the uncertain source of the trials effect discussed earlier, the data in Tables 3-9 and 3-10, and in Figures 3-3 and 3-4 represent averages across all six trials.)

Further analyses indicated that the main effect of experience with masks approached significance for the no mask condition ( $F = 2.66, p < .10$ ) and for the original mask condition ( $F = 2.48, p < .10$ ). A review of Figure 3-3 indicates that these differences appear to lie between the intermediate and high experience group for the no mask condition, and between the low and high experience groups for the original mask condition. It should be noted that

TABLE 3-7.  
 ANALYSIS OF VARIANCE SUMMARY TABLE FOR MISRECOGNITIONS  
 WITH MASK EXPERIENCE AS THE BETWEEN-GROUP VARIABLE

Source of Variance	df	MS	F
Experience (E)	2	1.33	7.37*
Error	15	0.18	-
Masks Condition (M)	2	1.01	10.39*
M x E	4	0.16	1.62
Error	30	0.09	-
Trials (T)	5	0.05	2.94*
T x E	10	0.01	0.60
Error	75	0.02	-
M x T	10	0.01	0.59
M x T x E	20	0.01	0.54
Error	150	0.02	-

\* p < .01

TABLE 3-8.  
ANALYSIS OF VARIANCE SUMMARY TABLE FOR MISRECOGNITIONS  
WITH MICROPHONE EXPERIENCE AS THE BETWEEN-GROUP VARIABLE

Source of Variance	df	MS	F
Experience (E)	1	2.05	9.91*
Error	16	0.20	-
Mask Condition (M)	2	1.42	15.12*
M x E	2	0.28	3.00
Error	32	0.09	-
Trials (T)	5	0.05	3.25*
T x E	5	0.01	0.50
Error	80	0.02	-
M x T	10	0.02	0.78
M x T x E	10	0.01	0.67
Error	160	0.02	-

\* p < .01

TABLE 3-9.  
 MEAN MISRECOGNITION ERROR RATES (IN PERCENT)  
 FOR LEVELS OF MASK EXPERIENCE BY MASK CONDITIONS

MASK CONDITION				
EXPERIENCE LEVEL	NO MASK	ORIGINAL MASK	SHURE MASK	$\bar{x}$ EXPERIENCE
Low	1.60	7.02	7.31	5.31
Intermediate	2.00	3.17	2.75	2.64
High	0.42	2.39	1.64	1.48
$\bar{x}$ MASKS	1.34	4.19	3.90	GRAND $\bar{x}$ = 3.14

TABLE 3-10.

MEAN MISRECOGNITION ERROR RATES (IN PERCENT)  
FOR LEVELS OF MICROPHONE EXPERIENCE BY MASK CONDITIONS.

## MASK CONDITION

EXPERIENCE LEVEL	NO MASK	ORIGINAL MASK	SHURE MASK	$\bar{X}$ EXPERIENCE
Low	1.54	6.41	7.06	5.00
High	1.07	2.83	1.76	1.89
$\bar{X}$ MASKS	1.30	4.62	4.41	GRAND $\bar{X}$ = 3.44

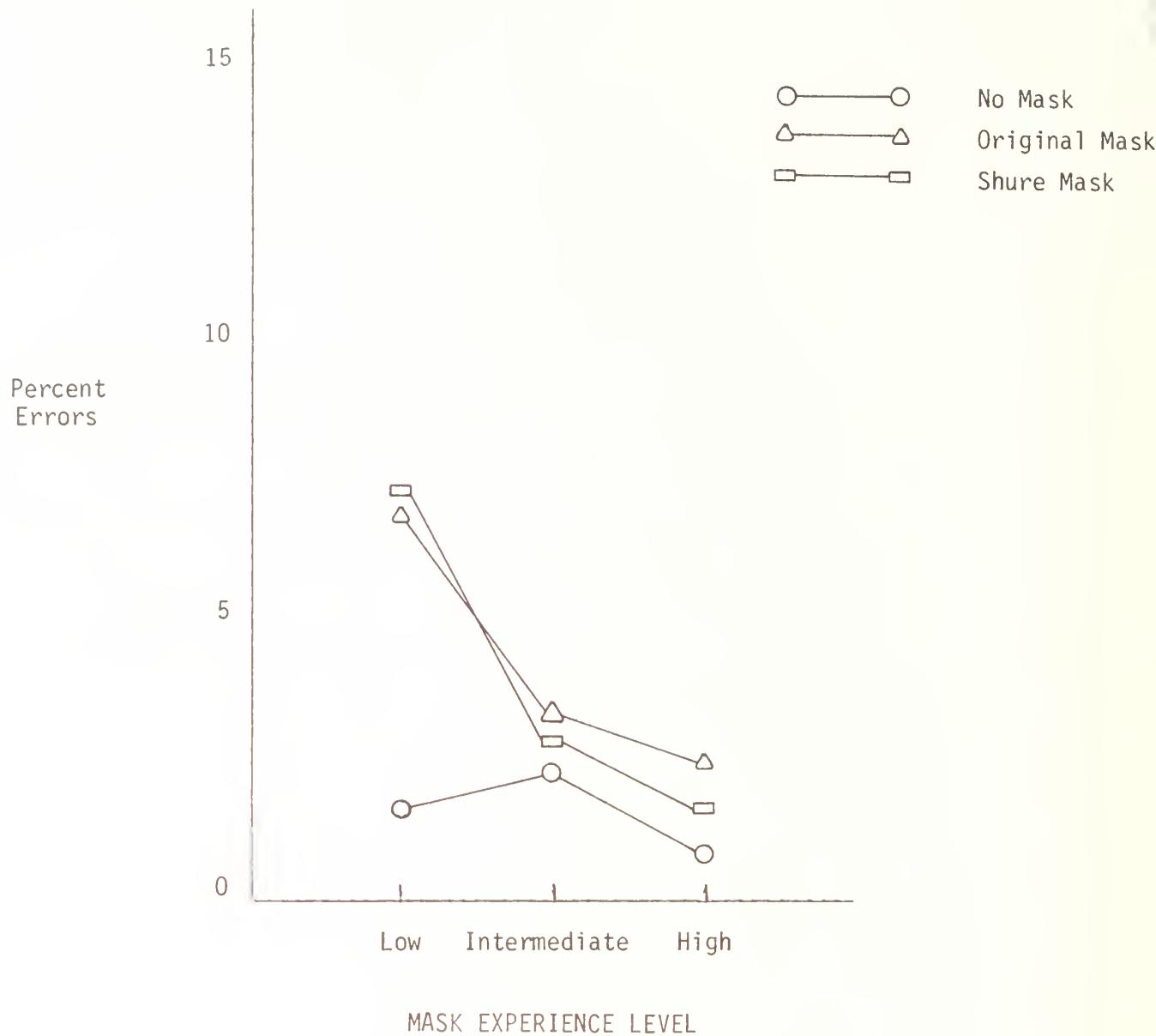
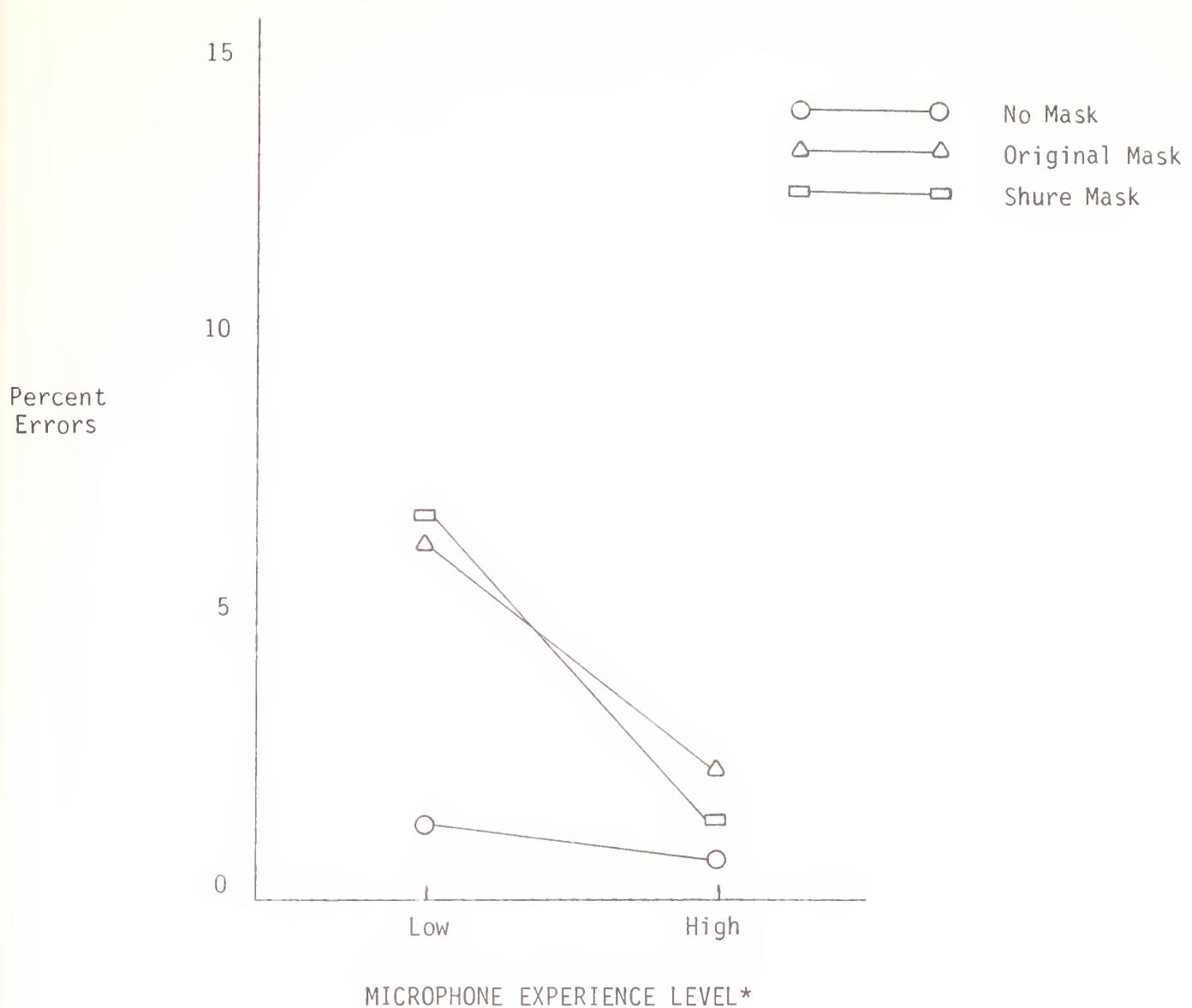


FIGURE 3-3.  
MISRECOGNITION ERROR RATES BY LEVELS OF MASK EXPERIENCE BY MASK CONDITIONS



\*There were no subjects of intermediate experience level with microphones.

FIGURE 3-4.  
MISRECOGNITION ERROR RATES BY LEVELS OF MICROPHONE EXPERIENCE BY MASK CONDITIONS.

even though this main effect is not significant at conventional statistical levels, the trend is in the expected direction and may be of practical (if not statistical) significance. The main effect of mask experience was statistically significant in the shure mask condition ( $F = 4.67$ ;  $p < .05$ ), and a Scheffé test indicated that the significant differences occurred between the low and high experience groups.

With regard to the main effect of experience with microphones, analyses performed on the experience levels for each mask condition indicated that the difference between the high and low experience groups (the only levels of experience for the microphone variable) was not significant under the no mask condition; under the original mask condition, this difference approached significance ( $F = 3.26$ ;  $p < .08$ ); and under the shure mask condition, the difference between high and low experience groups was highly significant ( $F = 10.19$ ;  $p < .01$ ).

A review of Figure 3-4 suggests that an interaction between mask condition and experience with microphones exists. This interaction approached significance ( $F = 3.00$ ;  $p < .06$ ), and suggests that the experience one had with microphones had more of a beneficial effect on error rates from the shure mask than it did on error rates from the original mask.

To determine whether the differences between mask groups were significant at each experience level, a series of one-way analyses of variance was performed on the misrecognition data using mask condition as the between groups variable. (Mean misrecognitions are those already reported in Table 3-9 for mask experience and 3-10 for microphone experience.)

For *mask* experience, the results were as follows: Significant differences were found between mask conditions for the low ( $F = 3.95$ ;  $p < .05$ ) and high ( $F = 5.55$ ;  $p < .05$ ) experience groups. Scheffé tests indicated that these

differences lie between the no mask and both original and shure mask conditions for the low experience group, and between the no mask and original mask conditions for the high experience group.

For *microphone* experience, significant differences were found between mask conditions for the low ( $F = 4.36$ ;  $p < .05$ ) and high ( $F = 3.47$ ;  $p < .05$ ) experience groups. Scheffé tests indicated that these differences lie between the no mask and shure mask conditions for the low experience group, and between the no mask and original mask conditions for the high experience group.



#### 4. DISCUSSION

Having presented the results of the present study, some implications of those results are now discussed.

##### 4.1 Total Errors

It is apparent that errors do increase when using voice technology under masked conditions. Table 3-2 showed an overall increase of roughly 3.5 percent between the no mask and (the average of) the original and Shure mask conditions. Viewing these data from the positive perspective, the no mask condition produced a total accuracy rate of 98.2 percent, which corroborates past research findings. The masked conditions produced an average accuracy rate of 94.7 percent (taken together) which, although (statistically) significantly worse than the no mask condition, is still quite impressive. One could argue that, depending on the particular application of "voice," this decrease in accuracy under masked conditions may not be *practically* significant.

Although the analyses conducted indicated a significant effect of trials, such that later trials seemed to produce a greater number of errors than earlier trials, this effect was restricted to the original mask condition, as shown in Figure 3-1. It is an interesting result, however, in that it is counter-intuitive; one would think that with practice, the error rate over trials should *decrease*. Several explanations are possible: First, it is entirely possible that 6 trials were not enough to display the performance improvement of a classical practice effect. More likely, however, is the explanation given previously, i.e., that the deterioration over trials is not a systematic but rather a spurious result. This is supported by the apparent absence of that effect for all but the original mask condition; if practice were a systematic effect, it should have

occurred under both mask conditions. As is suggested by the results of the experience variables tested, prolonged practice may in fact have a beneficial effect on overall performance with the "voice" system. Further research should investigate the effects of practice using a larger number of trials.

#### 4.2 Nonrecognitions

In general, there were no significant effects of any of the independent variables on nonrecognitions. That is, speaking into either the original or the Shure stenomasks did not appear to have any effect on the number of "beeps" or rejections emitted by the "voice" system. This is an encouraging finding in that it indicates an almost equivalent error rate for nonrecognitions across all mask conditions (see Table 3-3). Additionally, it should be noted that the highest nonrecognition rate (averaging across trials) for any of the mask conditions was approximately eight tenths of one percent (or a 99.2 percent accuracy rate). Thus, with regard to nonrecognitions, there should be no appreciable performance decrement when using masks with voice recognition equipment.

#### 4.3 Misrecognitions

The results for analyses of misrecognitions essentially parallel those for total errors. That is, mask condition did significantly affect performance such that more misrecognition errors were made while subjects spoke into masks. Essentially, both mask conditions appeared to contribute almost equally to the performance decrement.

A review of Table 3-5 shows, however, that the highest error rate (averaging over trials) was 4.62 percent (an accuracy rate of approximately 95.4 percent). Again, the accuracy rate for the no mask condition was impressive (98.7 percent), as found in past research.

The trials effect noted takes the same form as that noted in the analysis of total error rates, and the explanation given in section 4.1 applies here as well. Again, it is important to note that although the performance decrement displayed by subjects under masked conditions was statistically significant, the particular application of the voice system would probably determine whether or not this decrement has practical significance; there are no doubt quite a number of applications in which a 95.4 percent accuracy rate under masked conditions would be quite acceptable.

The performance decrement under masked conditions is perceived by the author (and by the researchers who were involved in conducting the study) to have been attributable in large part to subject's breathing into the stenomask between utterances. Apparently, the breaths taken with the masks in place resulted in misrecognition errors, as opposed to nonrecognition errors. Although subjects were instructed to remove the hand-held stenomask when they needed a breath (or to cut the circuit between the mask and the T600), some subjects still breathed into the masks, resulting in the T600 interpreting a breath as a spoken input. As will be discussed next, it is felt that this behavior could be largely eliminated, and error rates reduced markedly, by training subjects in how to speak into masks.

#### 4.4 Experience with Masks and Microphones

Significant and sizeable negative correlations were found between misrecognition error scores and subject's ratings of their experience with masks and microphones (see Table 3-6). Although not all significant, the direction of all the correlation coefficients presented in Table 3-6 suggests that the greater the amount of experience an individual has with speaking into masks and/or microphones, the lower the misrecognition error rates.

Further analyses (as described in section 3.5) showed that the experience effect was highly significant and (although not all differences between groups were statistically significant), Figures 3-3 and 3-4 show that the highly experienced subjects made far fewer errors (under masked conditions) than those subjects of low experience levels.

Tables 3-9 and 3-10 indicate that experience with masks and microphones had a somewhat beneficial effect even on performance under no mask conditions. Differences expressed in accuracy (instead of error) rates show that experience using either masks or microphones increased accuracy roughly from 93 to 98 percent. Although statistically significant differences still existed between several pairs of mask conditions even at high experience levels, these differences are likely to be insignificant for practical intents and purposes; an accuracy rate of roughly 97 percent in the worst case for highly experienced subjects is, again, rather impressive.

It is also important to note that the explanation given for misrecognition errors coming as a result of breathing into the masks receives considerable support from the findings regarding experience levels. It is clear that a major emphasis in pilot or communication training, for example, is placed on proper enunciation and control of implosions of consonants and other breath-control parameters. It follows, therefore, that those subjects experienced in the use of masks or microphones would have better control of these parameters, and would therefore perform better with regard to misrecognition errors. (Note also that although most correlations on the nonrecognition part of Table 3-6 were not statistically significant, the overall trend is for experience to be negatively correlated with nonrecognition errors).

## 5. CONCLUSIONS

The results of the present study are, in a word, encouraging. It is apparent that although using a stenographer's mask does contribute to an increase in the percent of misrecognition errors made, this increase in errors may be mitigated to a large extent by experience with speaking into masks or microphones. This leads the author to suggest that, with appropriate training, "masked" speakers could achieve an accuracy rate comparable to "unmasked" speakers using currently available voice recognition equipment. This opens the door to the potentially successful use of voice technology in many types of tactical and C<sup>3</sup> applications. In fact, research is now underway to determine the effectiveness of voice recognition equipment in situations where users are required to wear protective (gas) masks. What remains to be determined is the exact nature and costs of training "voice" users under various conditions, and the potential benefits of such training.



## 6. REFERENCES

Brown, M.G., Engelman, L., Frame, J.W., Hill, M.A., Jenurich, R.I., and Toporek, J.D. *BMDP Statistical Software 1981*, Los Angeles: University of California Press, 1981.

Bruning, J.L., and Kintz, B.L. *Computational Handbook of Statistics (2nd ed.)*, Glenview, Illinois: Scott, Foresman and Co., 1977.

Elster, R.S. The effects of certain background noises on the performance of a voice recognition system. Naval Postgraduate School, Monterey, California, Report No NP554-80-010, September, 1980.

Neter, J., and Wasserman, W. *Applied Linear Statistical Models*, Homewood, Illinois: Richard D. Irwin, Inc., 1974.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., and Bent, D.H. *Statistical Packages for the Social Sciences (2nd ed.)*, New York: McGraw-Hill, 1975.

Poock, G.K. Experiments with voice input for command and control: Using voice input to operate a distributed computer network. Naval Postgraduate School, Monterey, California, Report No. NP555-80-016, April, 1980.



APPENDIX A

LIST OF UTTERANCES



<u>WORD #</u>	<u>UTTERANCE</u>
0	ONE
1	TWO
2	YANKEE
3	AIR ROUTES
4	GARY POOCK
5	LOAD THE GANN
6	CARRIAGE RETURN
7	LOAD THE SERVER
8	IRAN
9	JAPAN
10	SWEDEN
11	EUROPE
12	LOGIN POOCK
13	LEVEL TWO
14	ACCAT TITLE
15	STRAIT OF HORMUZ
16	LOAD GLD3
17	CONNECT TO CHARLIE
18	POOCK NPS PASSWORD
19	CHANGE DIRECTORY TO HUNTER
20	THREE
21	FOUR
22	LOGOUT
23	GRAPHICS
24	RED SPHERE
25	STEAM PLANT

<u>WORD #</u>	<u>UTTERANCE</u>
26	ZERO
27	SEVEN
28	NOVEMBER
29	MOVE IT DOWN
30	USE THAT ONE
31	SPIROGRAPH
32	CAPTAIN EBBERT
33	CLOSE OUT CHARLIE
34	UP IN DETAIL
35	UNITED STATES
36	LEVEL TWO VIEWER
37	NORTH ATLANTIC MAP
38	GENISCO ZERO PARAMETERS
39	MEDITERRANEAN MAP
40	FIVE
41	SIX
42	ALPHA
43	BRAVO
44	CHARLIE
45	DELTA
46	ECHO
47	FOXTROT
48	JULIETT
49	ROMEO
50	MOVE IT LEFT

<u>WORD #</u>	<u>UTTERANCE</u>
51	SIERRA
52	SAN FRANCISCO
53	APPLICATION
54	ENGINEERING
55	HUMAN FACTORS
56	VOICE TECHNOLOGY
57	CENTRAL EXPRESSWAY
58	RUSSIAN VERSION OF HORMUZ
59	FILE TRANSFER PROTOCOL
60	EIGHT
61	NINE
62	HOTEL
63	INDIA
64	KILO
65	LIMA
66	OSCAR
67	POPPA
68	MOVE IT RIGHT
69	UNIFORM
70	VIETNAM
71	KOREA
72	ADVISORY
73	INTERACTIVE
74	BUSINESS MEETING
75	CONTINUOUS

<u>WORD #</u>	<u>UTTERANCE</u>
76	SPEECH RECOGNITION
77	CONTINUOUS SPEECH
78	EFFICIENT TRANSMISSION
79	SYSTEM INTEGRATION
80	GOLF
81	MIKE
82	QUEBEC
83	TANGO
84	VICTOR
85	WHISKEY
86	XRAY
87	ZULU
88	MOVE IT UP
89	BANGLADESH
90	TOKYO
91	HOLLISTER
92	DOWN IN DETAIL
93	CORPORATION
94	CRITERIA
95	ADVANTAGES
96	SUITABILITY
97	RADIOLOGY
98	IDENTIFICATION
99	AUTOMATIC RECOGNITION

APPENDIX B  
QUESTIONNAIRE



---

NAME

---

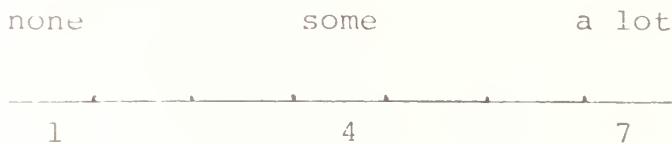
SUBJECT #

ON THE FOLLOWING PAGES YOU WILL FIND  
SEVERAL QUESTIONS/STATEMENTS DESIGNED TO  
GET YOUR REACTIONS TO USING VOICE RECOG-  
NITION EQUIPMENT. ALSO, THERE ARE  
QUESTIONS REGARDING YOUR EXPERIENCE WITH  
VARIOUS INPUT DEVICES.

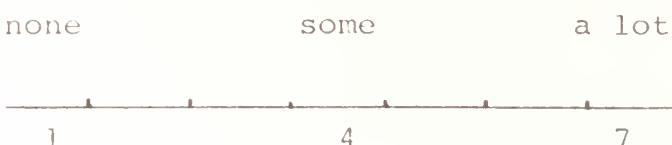
PLEASE RESPOND TRUTHFULLY, AND CHECK YOUR  
QUESTIONNAIRE AFTER COMPLETION TO MAKE SURE  
YOU'VE ANSWERED ALL THE ITEMS.

THANK YOU FOR YOUR COOPERATION AND PARTICIPATION  
IN THIS EXPERIMENT.

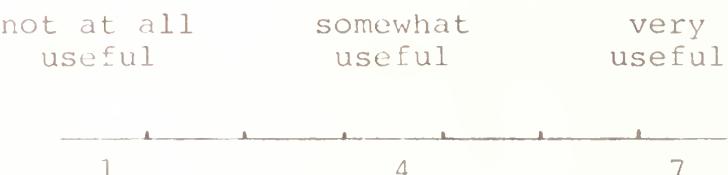
HOW MUCH EXPERIENCE HAVE YOU HAD IN USING MASKS (NOT INCLUDING THIS EXPERIMENT)?



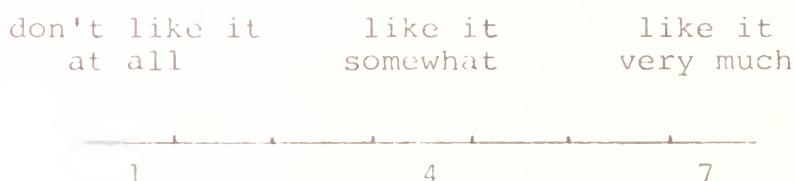
HOW MUCH EXPERIENCE HAVE YOU HAD IN SPEAKING INTO MICROPHONES  
(NOT INCLUDING THIS EXPERIMENT).



## HOW USEFUL DO YOU THINK VOICE RECOGNITION EQUIPMENT REALLY IS?



HOW MUCH DO YOU LIKE VOICE RECOGNITION EQUIPMENT?



PLEASE INDICATE THE EXTENT TO WHICH YOU AGREE OR DISAGREE WITH THE FOLLOWING STATEMENTS:

"I WOULD DO BETTER WITH VOICE EQUIPMENT IF I DIDN'T SEE OR HEAR WHEN I'VE MADE AN ERROR."

disagree      neither agree      agree  
strongly      nor disagree      strongly



"MAKING ERRORS WHEN USING VOICE EQUIPMENT IS FRUSTRATING."

disagree      neither agree      agree  
strongly      nor disagree      strongly



"I FEEL PRESSURED WHEN USING VOICE EQUIPMENT."

disagree      neither agree      agree  
strongly      nor disagree      strongly



"VOICE EQUIPMENT IS TOO HARD TO USE."

disagree      neither agree      agree  
strongly      nor disagree      strongly



"VOICE EQUIPMENT IS IMPRACTICAL."

disagree      neither agree      agree  
strongly      nor disagree      strongly



DISTRIBUTION LIST

No. of Copies

Library, Code 0142 Naval Postgraduate School Monterey, CA 93940	4
Dean of Research Code 012A Naval Postgraduate School Monterey, CA 93940	1
Library, Code 55 Naval Postgraduate School Monterey, CA 93940	1
Professor G. K. Poock Code 55Pk Naval Postgraduate School Monterey, CA 93940	100

U203692



5 6853 01069201 5

1203692